다양한 활성 함수를 사용하는 신경회로망의 구성

*이충열, 박철훈
한국과학기술원 전자전산학부
e-mail : chungyeol.lee@gmail.com, chpark@kaist.ac.kr

Neural Networks with Mixed Activation Functions

*Chung-Yeol Lee, Cheol Hoon Park
School of Electrical Engineering and Computer Science
Korea Advanced Institute of Science and Technology

Abstract

When we apply the neural networks to applications, we need to select proper architecture of the network and the activation function of the network is one of most important characteristics. In this research, we propose a method to make a network using multiple activation functions. The performance of the proposed method is investigated through the computer simulations on various regression problems.

I. Introduction

The Multi Layer Perceptron (MLP) and Radial Function Network (RBF) are most famous ones among them but they have quite different characteristics. The MLP uses a sigmoid function generally as its activation function while the most RBF uses Gaussian. Each neuron in MLP have global characteristic, i.e., change of a weight belong to a neuron makes change output of it in wide range. On the other hand, a neuron in RBF have local characteristic that makes changes only in an activation region of a neuron when we change a weight of it.

When we use an ANN, we should choose proper type of network from the property of a given problem. However, we need both global and local characteristics to obtain good performance in general. In this paper, we propose a method to build a network using both two activation functions of MLP and RBF to solve regression problem. Using both neurons, we can obtain more reliable performance on various problems. Also, the result networks are more efficient having smaller number of neurons than a network with a single activation function. The detailed algorithm is given in the following section. Section 3 shows the computer simulation results. Finally, conclusion is made in Section 4.

II. A Proposed Algorithm

For given n input and target pairs, $\{(\mathbf{x}_i, t_i)\}_{i=1}^{n}$,, we want to obtain an ANN hich satisfies the Root Mean Squared Error (RMSE) for training samples is smaller than a given goal. The RMSE is as follows:

$$E=\sqrt{\frac{1}{n}\sum(t_i-\mathrm{net}(\mathbf{x}_i;\mathbf{w}))^2} \quad (1)$$

where $\mathrm{net}(\mathbf{x};\mathbf{w})$ is an output of network having parameters $\mathbf{w}$ and input $\boldsymbol{x}_i$.

When we try to construct a network by selecting the best neuron at the moment among the given activation functions, which gives the largest error minimization, there is a tendency the local property is preferred. The sigmoid function sometimes needs multiple neurons to fit the function properly, but when they are fitted they work better than local neurons. To avoid this problem, we take an approach to prune the neurons from the trained network.

Let there is a given initial number of neurons. To get a network with two different activation functions, we first construct each network which has a half of the given initial number of neurons of which activation function are tangent sigmoid and Gaussian function, respectively. For the MLP part, we initialize the network weights to be uniformly distributed between ± 0.05 and for RBF, we use OLS [1] algorithm. Then, each network is trained using Levenberg-Marquadt (LM) [2] algorithm for given training samples. After training each network, we merge them to a network. Now, the network have a given initial number of neurons with mixed kernel. Finally, we can obtain an efficient network by applying Optimal Brain Surgeon (OBS) [3] pruning algorithm.

Ⅲ. Experiments

We solve the five different regression problems in Fig. 1. All of them are 2-dimensional nonlinear functions and have a domain $[-1,1]^2$. Among the five functions, function (b) and (e) are well fitted with sigmoid and the function (c) is well fitted with Gaussian. Function (a) and (d) have both of global and local properties.

We compared the proposed algorithm to the 2-layer MLP and RBF networks. The networks are initialized, trained and then pruned in the same way in proposed method. The only difference is that the network has a number of given initial number of neurons with one kind of activation function. We set the initial number of neurons to 20 for function (a), (c), and (e) and 26 for (b) and (d) and use 0.02 as the training goal in RMSE. To compare the performance, we repeated the experiments ten times for each function with different initial weights and measure the success rate. We also measure the mean value of training and generalization error in RMSE and the number of neurons for successful cases. The result are in Table 1. The proposed algorithm shows better performance than others. The proposed one always succeed to achieve the given training goal. In the view point of efficiency, we can obtain smaller size network with proposed method for function (a) and (d) which have both of global and local properties. For the other functions, the proposed algorithm gives us a reasonable performance compared to the others. Even though function (c) can be fitted well with Gaussian, the results shows that the using sigmoid and Gaussian functions together improve the success rate much.

Table 1: The result of experiments

Function (a)	Mixed	MLP	RBF
Success	10	10	7
# of neurons	11.9	12.1	13.43
Training	0.0155	0.0173	0.0178
Generalization	0.0151	0.0171	0.0162
Function (b)	**Mixed**	**MLP**	**RBF**
Success	10	10	1
# of neurons	18.8	16.5	21.0
Training	0.0154	0.0166	0.0112
Generalization	0.0199	0.0210	0.0154
Function (c)	**Mixed**	**MLP**	**RBF**
Success	10	0	4
# of neurons	8.4	-	6.5
Training	0.0107	-	0.0082
Generalization	0.0170	-	0.0154
Function (d)	**Mixed**	**MLP**	**RBF**
Success	10	10	0
# of neurons	18.0	23.1	-
Training	0.0175	0.0188	-
Generalization	0.0230	0.0340	-
Function (e)	**Mixed**	**MLP**	**RBF**
Success	10	10	1
# of neurons	10.1	7.3	19.0
Training	0.0140	0.0136	0.0174
Generalization	0.0148	0.0137	0.0252

Ⅳ. Conclusion

We proposed a method to make an network uses both tangent sigmoid and Gaussian functions for a regression problem. Using both activation functions, the network shows learn the training samples efficiently. In the computer simulation, the proposed one shows better performance than MLP and RBF networks.

References

[1] S. Chen, C. F. N. Cowan, and P.~M. Grant, "Orthogonal least squares learning algorithm for radial basis function networks," *IEEE Transactions on Neural Networks*, vol. 2, no. 2, pp. 302-309, Mar. 1991.

[2] C. M. Bishop, *Neural Networks for Pattern Recognition*. New York: Clarendon Press, 1995.

[3] B. Hassibi and D. G. Stork, "Second order derivatives for network pruning: Optimal brain surgeon," in *Advances in Neural Information Processing Systems*, S. J. Hanson, J. D. Cowan, and C.~L. Giles, Eds., vol. 5. Morgan Kaufmann, 1993, pp. 164-171.

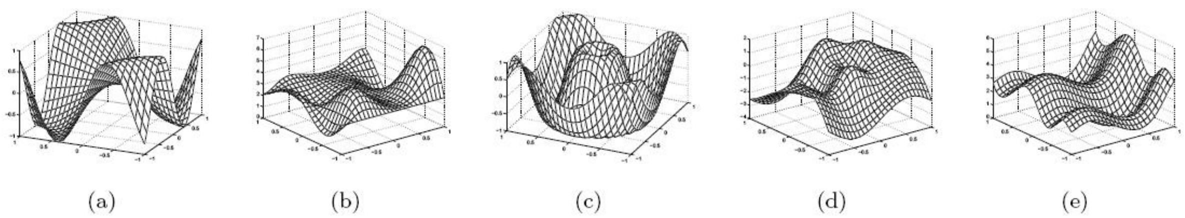

Figure 1. 2-Dim Regression Problems